DOCUMENT RESUME

ED 244 513 FL 014 305

AUTHOR Hahn, Sidney L.
TITLE MLA Scores as Predictors of Success in Student Teaching.
PUB DATE May 84
NOTE 23p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Evaluation Criteria; French; German; *Language Teachers; *Language Tests; *Predictor Variables; Rating Scales; Spanish; *Standardized Tests; *Student Teaching; *Teacher Effectiveness; Weighted Scores
IDENTIFIERS *MLA Cooperative Foreign Language Tests; University of Nebraska Lincoln

ABSTRACT

An investigation of the predictive power of the Modern Language Association (MLA) Cooperative Test Scores in French, German, and Spanish in relation to performance in student teaching is described, and criteria used by the raters to define successful performance are identified. Subjects were 103 students who completed foreign language teaching requirements during the years 1973-1981 at the University of Nebraska-Lincoln. Student teaching performance was rated by three university supervisors and three secondary school principals who screened the written narrative comments appearing in each student teacher's final credential. Findings supported a significant positive relationship between MLA scores and successful student teaching. A significant difference was found in the means of the two groups of raters, although the ratings were highly correlated and similar evaluative criteria had been developed. The difference appeared to be related to the weighting of identified criteria. (Author/MSE)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED244513

# MLA Scores as Predictors of Success in Student Teaching

by

Sidney L. Hahn

University of Nebraska-Lincoln

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Sidney L. Hahn

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL017305

# MLA Scores as Predictors of Success in Student Teaching

## ABSTRACT

This paper 1) describes an investigation of the predictive power of MLA Cooperative Test scores in French, German, and Spanish in relation to performance in student teaching, and 2) identifies the criteria used by the raters to define successful performance. Subjects were 103 students who completed foreign language teaching endorsements during the years 1973-1981 at the University of Nebraska-Lincoln. Student teaching performance was rated by three university supervisors and three secondary school principals who screened the written narrative comments appearing in each student teacher's final credential.

Findings supported a significant positive relationship between MLA scores and successful student teaching. A significant difference was found in the means of the two groups of raters, although the ratings were highly correlated, and similar evaluative criteria were developed. The difference appeared to be related to the weighting of identified criteria.

RATIONALE AND PURPOSE

Teachers in all subject areas are expected to have expertise and knowledge in their chosen discipline. In today's modern language classroom, the teacher's oral command of the target language is considered essential. Who would deny that teachers must have enough command of the grammatical structure of a language to be able to both model its correct use and explain its critical features. It seems logical to assume that such subject area competency would be one factor in successful student teaching, but how much control is needed? Is there a measurable relationship between foreign language competency and successful student teaching? Can MLA Cooperative Test scores be used in assessing the entry level language competence needed for successful student teaching?[1] Is one skill (i.e., listening, speaking, reading, writing) more indicative of success than another?

The literature revealed no prior research of the above questions. MLA scores have been routinely used by the investigator for purposes of advising and making practicum placement decisions. Course grades were the only criteria applied by UNL (Teachers College) in screening candidates for student teaching during the years 1973-1981. The purpose of the investigation was to compare MLA scores with successful student teaching as perceived by 1) school administrators (hiring

officials and 2) university supervisors (writers of evaluation summaries.)

DESIGN

A study was designed to investigate the predictive power of MLA Cooperative Listening, Speaking, Reading and Writing test scores in French, German, and Spanish in relation to performance in student teaching. Narrative comments on final evaluations were screened by one group of three university supervisors and one group of three secondary school administrators (principals). Student teaching success was rated on a scale of 1 (high) to 4 (low). The question was considered in relation to 1) student teachers of French, German, and Spanish as a combined group; 2) student teachers of Spanish; 3) student teachers of French; and 4) student teachers of German.

Related questions studied were: 1) What criteria are identified by supervisors and principals in rating success as high or low on a scale of 1 to 4? 2) Are perceptions of successful student teaching the same when written comments on final evaluations are screened by supervisors and principals?

METHODOLOGY

Subjects were 103 students who completed teaching endorsements in the areas of French, German, and Spanish at the University of Nebraska from 1973-1981. MLA test scores (Form MA) were routinely obtained during the methods block (Phase III) of the teacher preparation program which precedes student teaching (Phase IV).

(M-Forms of the MLA Ccoperative Tests were selected in 1972 by the Advisory Council of the NFLA and Nebraska State Department of Education for use by all Nebraska colleges; the 50th percentile was recommended as a minimum graduation requirement for FL majors.)

A training session was held for each group of raters, 1) three university supervisors, and 2) three secondary (junior or senior high school) principals. Each rater independently judged the performance of 28 graduates for whom no MLA scores were available. Raters saw only the narrative written comments from the final evaluation forms; no other information was given. The narrative comments were those agreed upon during the regular final evaluation conference of the college supervisor with the supervising teacher and student teacher as summarized and submitted by the supervisor on the standard department form. (The form is retained as a permanent part of the teaching credential.)

Using the written evaluation comments as the basis for decisions, raters first separated the Ss into high and low groups, then subdivided them on a scale of (high) 1, 2, and (low) 3, 4. All cases which showed discrepant ratings were discussed and the discussions tape recorded. Raters were directed to attempt to reach consensus. If raters were unable to agree the majority opinion (two of three) prevailed for purposes of statistical analysis. The attempt to reach consensus served to isolate and identify those criteria raters considered important in giving

their ratings, thus adding information about the judgments made. This procedure is similar to that used in the final student teacher evaluation conference.

Each group, in a separate session, then rated the performance of the 103 Ss using the above procedure. Each group of raters was also interviewed by the investigator at the conclusion of the formal evaluation sessions in regard to process and criteria for decision making.

DATA ANALYSIS

A correlation matrix (Pearson Correlation) was obtained for the variables Listening, Speaking, Reading, Writing, supervisor ratings and administrator ratings by combined language groups and by separate language.[2]

Multiple regression analysis was used to determine the predictive power of the set of independent variables (Listening, Speaking, Reading, Writing) in the following relationships: 1) principals' ratings of successful performance of 103 student teachers of French, German, and Spanish; 2) ratings of successful performance of 45 student teachers of Spanish; 3) ratings of successful performance of 26 student teachers of French; and 4) ratings of successful performance of 32 student teachers of German. The significance of t was obtained for the independent variables. The principals' ratings provided the dependent variable for each multiple regression analysis. The investigator chose not to run separate regression analyses for the following

reasons: 1) Supervisor and principal ratings were highly correlated (.826); 2) very similar criteria for ratings were identified by both rater groups including high emphasis given to subject area competency, the major factor under study, and 3) chi square analysis of rater criteria showed no significance of difference within factor clusters between the two groups of raters.

A t test for correlated means was used to compare paired means of the two groups of raters in regard to similarity or divergence of perceptions of student teaching performance.[2]

Criteria developed by the raters in analyzing successful student teaching performance were identified by tabulation of the factors specified during the tape-recorded discussions. The rater-identified factors were then grouped into clusters as shown in Table V. Related data obtained by interview were compared for consistency, and will be discussed in relation to the findings.

## RESULTS AND DISCUSSION

Positive MLA score relationships, well below the .05 level of significance, were found for the combined groups of languages (P = .0145), for Spanish (P = .0138), and for German (P = .0270). Subtests shown to be significant discriminating variables were: Reading, combined languages (t = .0391); Listening, Spanish (t = .0087); and Writing, French (t = .0391). Tables 1 and 2 display these data. No single variable was significantly predominant

Insert Tables 1 and 2 about here

in German although Reading was shown to be of relatively higher importance in the Beta statistic (Table 2). All variables were correlated, as shown in Table 3.

Insert Table 3 about here

A significant difference (<.01) was found in perceptions of successful student teaching between the two groups of raters (See Table 4).

Insert Table 4 about here

The two groups of raters identified similar evaluation criteria in assigning ratings of success. Table 5 shows the clusters of factors which influenced raters' decisions, either positively or negatively. The distribution of single factors varied in the frequency with which each was named as influential. Chi square analysis failed to show significance of difference however, perhaps attributable to the fact that some cells contained few entries, perhaps only to effects of random differences in individual standards by members of each group. (The latter hypothesis can be further tested in a follow-up rater by group by subject analysis of variance).

Insert Table 5 about here

MLA scores for French, German, and Spanish were shown to be predictive of successful performance in student teaching when rated as described. The significant variable for the combined language groups was Reading. (The means of adjusted mid-percentile scores for the Ss on the Reading variable were: Combined languages (N = 103) $\bar{X}$ = 81.00; Spanish (N = 45) $\bar{X}$ = 78.86; French (N = 26) $\bar{X}$ = 82.67; German (N = 32) $\bar{X}$ = 82.66.)

When each language was considered separately, predictive significance of the MLA scores was found for Spanish and for German. Significant variables were found to be: Spanish, Listening (N = 45) $\bar{X}$ = 82.96; French, Writing (N = 26) $\bar{X}$ = 73.56; German, none.

The statistical findings were also consistent with the principals' and supervisors' stated concern for subject area competency as a factor. Language competency was specifically named 22 times by the group of principals in making judgments, that total being highest of any single factor identified. As shown in Table V (Clusters of Factors), only interpersonal abilities and skills ranked as more important. (The latter category contained several factors which were considered similar enough to be combined.) When asked: "Did statements related to

French/German/Spanish language control affect your ratings?", the answers of both groups were all affirmative.

In interpreting the statistical findings it should be noted that MLA scores obtained by the Ss were skewed in a direction higher than the norms on which the scores are based. (Combined language means were: 1) listening, 85.79; 2) speaking, 77.58; 3) reading, 81.00; 4) writing, 82.59). This was to be expected since original norms were based on two years of college language study, whereas the majority of the Ss in this study were either second semester juniors or first semester seniors nearing completion of requirements for the foreign language major. A few had completed the baccalaureate degree and had returned to receive teaching certification. In this regard, an assumption of the study was that the Ss represented a normal distribution of University of Nebraska students seeking foreign language teaching endorsement.

The finding of significant difference in the means of the two groups of raters raises such questions as 1) why, and 2) which factors contributed to the difference? The finding of a significant difference is consistent with that reported by Reinhartz and Beach in a related study, but with one major discrepancy.[3] That study also reported a significant difference between supervisor and principal ratings, but the supervisors consistently rated teachers higher than did the principals when using the same criteria. The supervisors in the current (MLA test) study, as a group, had consistently lower mean ratings than

the principals for both separate and combined language groups. The difference may be related to self selection (identification) of criteria on the part of each group of raters in the later study as opposed to being given criteria in the former study, or the fact that the original evaluation comments were formulated cooperatively by supervisor and supervising teachers rather than principals. The frequency distribution shown in Table V, although statistically inconclusive, offers one intriguing discrepancy, that being the high incidence of personal qualities identified as influential by the supervisors. These supervisors were also more concerned with the wording of the recommendations and with the inclusion of a specific statement of recommendation (or lack thereof).

IMPLICATIONS

The continued use of the MLA Cooperative Tests (and particularly of the Reading subtest) would seem to be indicated as predictive of success in student teaching given similar conditions. The Reading/Listening and Writing/Speaking scores were generally shown to be more highly correlated (Table III), and all variables were positively correlated with administrator ratings (Multiple R = .0343); thus administering only the reading and writing subtests could provide a rough indication of relative Listening/Speaking control. Scoring may then be done largely by key, another time-saving factor. The speaking subtest did not prove to be of significant value. The MLA listening and speaking tests do not

provide a rating of interactive skills under circumstances of either real or simulated communication. Recently developed tests and procedures for scoring of oral communicative competence may provide future means of shedding light upon such relationships. Research should continue concerning the speaking skill in relation to teacher competency.

The conclusions cited are not meant to imply that factors other than language competence are either unimportant or uncorrelated with student teaching success. No attempt was made in this study to test either related factors or interaction effects. Items which emerged as candidates for further research are those clustered in Table 5.

Further analysis of the data gathered in this study concerning supervisor/administrator selection of criteria and idiosyncratic semantic interpretation of the narrative comments (together with implications for supervisor training) remain to be carried out and reported. Since principals routinely screen the evaluations prepared by supervisors in selecting candidates for interviewing, further study of semantic factors as influences could prove interesting and enlightening.

NOTES

[1]MLA-Cooperative Foreign Language Tests, Educational Testing Service, Cooperative Test Division, Princeton, New Jersey, 1965.

[2]Norman H. Nie, C. Hadlai Hull, Jean G. Jenkins, Karin Steinbrenner, & Dale H. Bent, Statistical Package for the Social Studies, second edition, (New York: McGraw-Hill Book Co., 1975, pp. 176-86, [Pearson Correlation]; pp. 320-60 [Multiple Regression]; pp. 249-74 [T-Test]).

[3]Judy Reinhartz & Don M. Beach, "A Comparison of Preservice and Inservice Ratings," Phi Delta Kappa, CEDR Quarterly, 15, 3 (1982) pp. 11-13.

TABLE 1

Multiple Regression Analysis for Combined Language Groups

| Variable | B | Beta | T | ANNOVA Source | DF | Sum of Squares | Mean Square | F |
|---|---|---|---|---|---|---|---|---|
| Write | 0.00594 | 0.14330 | 0.976 | Regression | 4 | 8.94708 | 2.23677 | 3.27575* |
| Read | -0.01633 | -0.28980 | -2.091* | Residual | 98 | 66.91700 | 0.68283 | |
| Speak | -0.84735D-04 | -0.00231 | -0.018 | | | | | |
| Listen | -0.00968 | -0.16479 | -1.169 | | | | | |

Multiple R = 0.34342

*P = 0.0145, significant for alpha = .05

*Reading, T = 0.0391, significant for alpha = .05

TABLE 2

Multiple Regression Analysis by Language Group

| Variable | B | Beta | T | ANNOVA Source | DF | Sum of Squares | Mean Square | F |
|---|---|---|---|---|---|---|---|---|
| SPANISH: | | | | | | | | |
| Write | 0.01154 | 0.30676 | 1.390 | Regression | 4 | 4.95339 | 1.23835 | 3.57733* |
| Listen | -0.02220 | -0.49907 | -2.758** | Residual | 40 | 13.84661 | 0.34617 | |
| Speak | 0.00302 | 0.08565 | -0.453 | | | | | |
| Read | -0.01483 | -0.29372 | -1.528 | | | | | |
| Multiple R = 0.51330 | | | | *P = 0.0138, significant for alpha = .05 | | | | |
| *Listening, T = 0.0087, significant for alpha = .01 | | | | | | | | |
| GERMAN: | | | | | | | | |
| Write | -0.01481 | -0.17918 | -0.596 | Regression | 4 | 10.69421 | 2.67355 | 3.24074* |
| Speak | -0.00931 | -0.18318 | -1.023 | Residual | 27 | 22.27454 | 0.82498 | |
| Listen | 0.00583 | 0.06631 | 0.258 | | | | | |
| Read | -0.02703 | -0.36083 | -1.048 | | | | | |
| Multiple R = 0.56954 | | | | *P = 0.0270, significant for alpha = .05 | | | | |

TABLE 2 (continued)

Multiple Regression Analysis by Language Group

| Variable | B | Beta | T | ANNOVA Source | DF | Sum of Squares | Mean Square | F |
|---|---|---|---|---|---|---|---|---|
| FRENCH: | | | | | | | | |
| Write | 0.03121 | 1.03681 | 2.114* | Regression | 4 | 4.75219 | 1.18005 | 1.57276 |
| Read | -0.01661 | -0.37098 | -1.115 | Residual | 21 | 15.86320 | 0.75539 | |
| Listen | -0.00898 | -0.16893 | -0.476 | | | | | |
| Speak | -0.02889 | -0.83617 | -1.858 | | | | | |

Multiple R = 0.48012 P = 0.2184

*Writing, T = 0.0466, significant for alpha = .05

TABLE 3

Pearson Correlation Coefficients for MLA Subtests by Language

| Combined Languages: | Listen | Speak | Read | Write |
|---|---|---|---|---|
| Listen | 1.000 | 0.505 | 0.684 | 0.612 |
| Speak | 0.505 | 1.000 | 0.402 | 0.650 |
| Read | 0.684 | 0.402 | 1.000 | 0.611 |
| Write | 0.612 | 0.650 | 0.611 | 1.000 |
| Spanish: Listen | 1.000 | 0.562 | 0.550 | 0.575 |
| Speak | 0.562 | 1.000 | 0.449 | 0.656 |
| Read | 0.550 | 0.449 | 1.000 | 0.678 |
| Write | 0.575 | 0.656 | 0.678 | 1.000 |
| French: Listen | 1.000 | 0.603 | 0.776 | 0.687 |
| Speak | 0.603 | 1.000 | 0.393 | 0.886 |
| Read | 0.776 | 0.393 | 1.000 | 0.578 |
| Write | 0.687 | 0.886 | 0.578 | 1.000 |
| German: Listen | 1.000 | 0.136 | 0.774 | 0.658 |
| Speak | 0.136 | 1.000 | 0.336 | 0.420 |
| Read | 0.774 | 0.336 | 1.000 | 0.835 |
| Write | 0.658 | 0.420 | 0.835 | 1.000 |

TABLE 4

T-Test of Differences in Paired Means of Ratings of Successful Student Teaching

| Rater Group | N | Mean | SD | Corr. | T | DF |
|---|---|---|---|---|---|---|
| Supervisors | 103 | 2.03 | 0.90 | 0.826 | 5.08* | 102 |
| Principals | 103 | 1.77 | 0.86 | | | |

*Significant <.01 level, 2 - Tail Prob.

TABLE 5

Rating Criteria Clusters Identified by Supervisors and Administrators

| Criteria | Absolute Frequency | | Percent | |
|---|---|---|---|---|
| | Adm. | Supv. | Adm. | Supv. |
| A. Teaching Skills | | | | |
| 1. Planning & Preparation | 18 | 17 | 10 | 8 |
| 2. Instruction & Evaluation | 20 | 21 | 11 | 10 |
| 3. Subject (FL) Control | 22 | 18 | 12 | 8 |
| TOTAL | 60 | 56 | 34 | 26 |
| B. Individual Qualities | | | | |
| 1. Personal Characteristics | 22 | 50 | 12 | 23+ |
| 2. Professional Characteristics | 27 | 33 | 15 | 15+ |
| 3. Interpersonal Abilities/Skills | 34 | 26 | 19 | 12 |
| TOTAL | 83 | 109 | 46 | 51 |
| C. Format of Narrative Comments | | | | |
| 1. wording/Flavor | 27 | 37 | 15 | 17 |
| 2. Consistency/Balance | 9 | 13 | 5 | 6 |
| TOTAL | 36 | 50 | 20 | 23 |
| GRAND TOTAL ALL CLUSTERS | 179 | 215 | 100 | 100 |